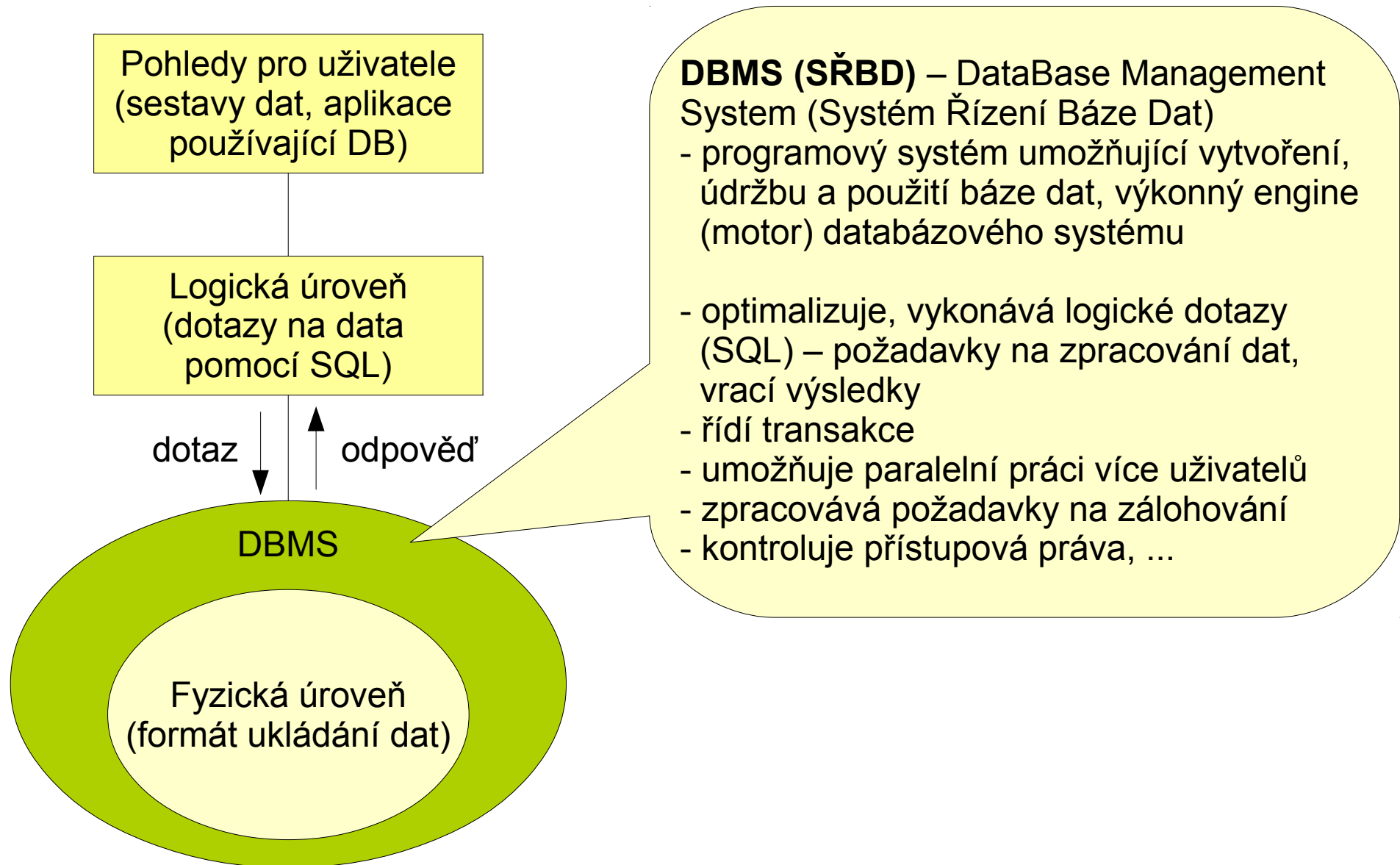


# Úvod do relačních databází

# Databáze a databázový systém

- **Databáze** – soubor informací/dat, báze dat (databanka), „černá skříňka s množstvím organizovaných informací“.
- **Databázový systém** = data + nástroje pro práci s daty.
- Nástroje pro práci s daty umí:
  - Vyhledávat, vkládat, aktualizovat, rušit data,
  - definovat strukturu dat,
  - definovat integritní omezení (vztahy mezi daty, omezení kladená na data),
  - spravovat uživatele databázového systému,
  - řídit transakce,
  - zálohovat.

# Database Management System



# Nezávislost dat

- **Fyzická nezávislost dat:**
  - Způsob fyzického uložení dat je oddělen od způsobu práce s daty (tj. od logické úrovně v podobě SQL).
- **Logická nezávislost dat:**
  - Hovoříme o ní, pokud změna logické struktury dat nenaruší fungování již existujících dotazů do databáze.

# Modely uspořádání dat v databázích

- **Hierarchický model** – data jsou organizována hierarchicky (do stromu), mezi tabulkami s daty je vztah rodič – potomek (potomků může být více),
- **Sít'ový model** – tabulky s daty jsou organizovány do sít'ové struktury, kde existují vztahy vlastník – člen (vlastník může mít více členů/složek, člen jednoho vlastníka),
- **Relační model** – data jsou organizována do n-tic, které se mohou v paměti kombinovat s dalšími n-ticemi, lze vyjádřit všechny typy vztahů (1:1, 1:N, N:1, M:N). Vazby mezi daty nejsou pevné, konkrétní provázání se provede až podle potřeby (na základě příkazu jazyka SQL).

# Relační model dat

- Představen roku 1969 dr. Edgarem F. Coddem z IBM.
- Založen na matematických základech: Teorii množin a predikátové logice.
- Data jsou organizována do n-tic (např. jméno, příjmení, věk konkrétního zaměstnance), jejichž členy (atributy) mohou být libovolně řazeny za sebou dle požadavků, n-tice stejné povahy lze skládat za sebe a řadit podle stanovených kritérií.
- Množinám n-tic se říká relace. V databázi jsou stanoveny pevné množiny n-tic (trvalé relace) – databázové tabulky.
- Data jsou uložena nezávisle na povaze databázové aplikace.
- S daty se manipuluje pomocí jazyka SQL, kterým určujeme, co chceme s daty dělat, nikoliv jak toho dosáhnout (neprocedurální jazyk).

# Základní pojmy relačních databází

- **Entita** – objekt reálného světa s vlastnostmi, které mohou být popsány v databázi (např. sousedův pes „Azor“).
- **Vazba (relationship)** – vztah mezi dvěma nebo více entitami (např. Azor „potrhal“ naše kočky Lízu a Mourka).
- **Entitní typ (sada entit)** – formální popis společných vlastností určité skupiny objektů (např. „pes“ obecně, s vlastnostmi pohlaví, plemeno, věk, počet zubů).
- **Atribut** – vlastnost entity určená názvem, typem a případně hodnotou (např. „počet zubů“ s typem „integer“, hodnotou 42).

# Základní pojmy relačních databází

- **Doména** – množina všech hodnot, kterých může atribut nabývat (např.  $\{0, 1, 2, \dots, 42\}$ ).
- **N-tice (tuple)** – n-tice atributů (s udaným názvem, typem a hodnotou), u kterých nezáleží na pořadí, (řádek v databázové tabulce, většinou popisuje jednu entitu).
- **Relace** – obecně množina n-tic. Může být trvalá (databázová tabulka), dočasná (množina n-tic aktuálně vytvořená v paměti při zpracování dotazu) nebo odvozená (určitý upravený pohled na trvalou relaci).






# Pojmy aplikované na tabulku

Entita – konkrétní stát  
(Trvalá) relace – tabulka countries

Název atributu

N-tice relace  
(řádek)

Záhlaví relace  
(schéma)

 COUNTRY_ID	 COUNTRY_NAME	 REGION_ID
1 AR	Argentina	2
2 AU	Australia	3
3 BE	Belgium	1
4 BR	Brazil	2
5 CA	Canada	2
6 CH	Switzerland	1
7 CN	China	3
8 DE	Germany	1
9 DK	Denmark	1
10 EG	Egypt	4
11 FR	France	1
12 HK	HongKong	3
13 IL	Israel	4
14 IN	India	3

Tělo relace  
(instance)

Hodnoty atributu  
COUNTRY\_ID.  
Všechny  
povolené  
hodnoty tvoří  
doménu

Atribut (sloupec)

# Pravidla pro tabulkovou reprezentaci dat

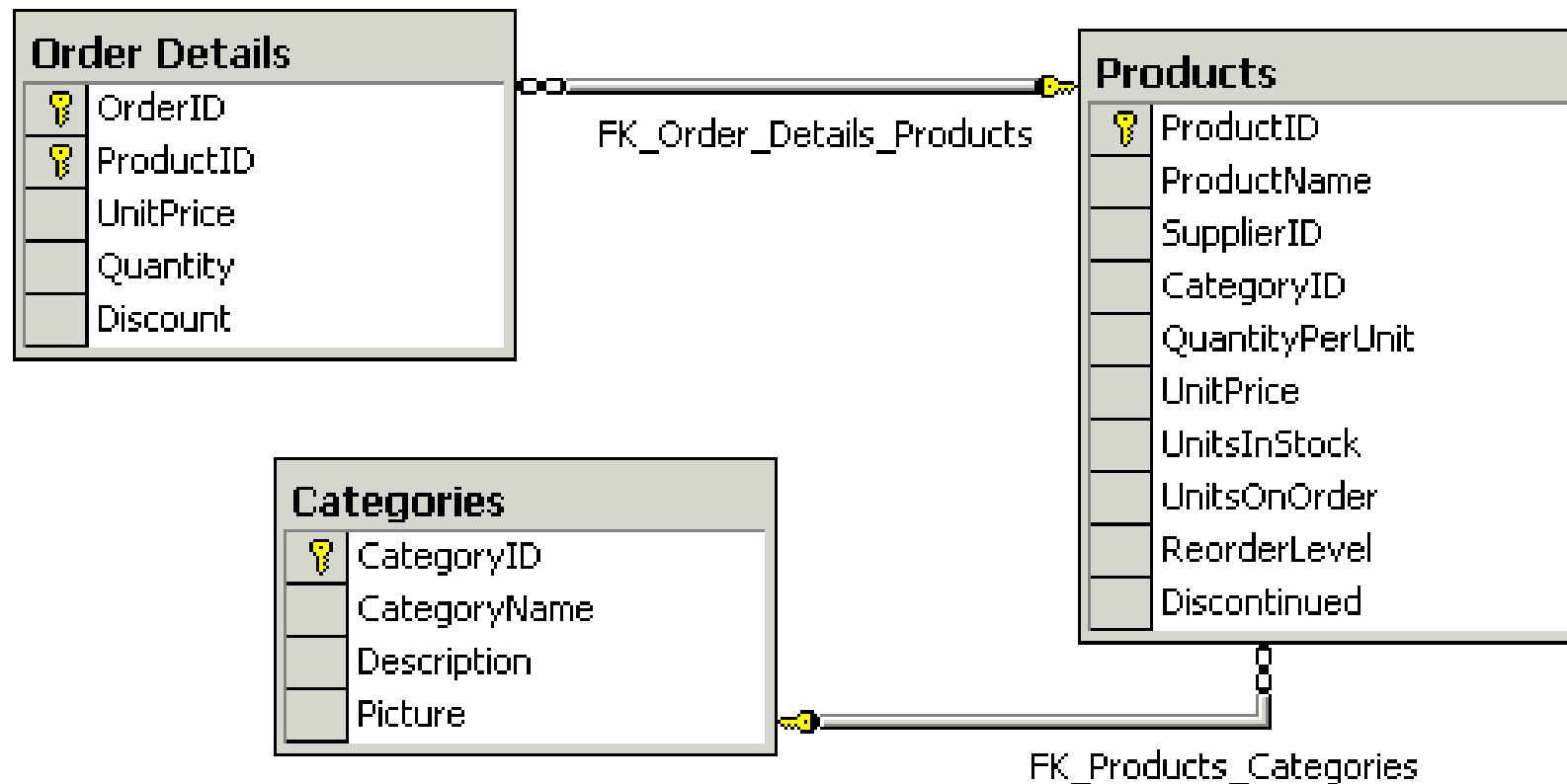
- Každý řádek odpovídá jedné n-tici relace,
- pořadí řádků není významné,
- pořadí sloupců není významné,
- žádné dva řádky neobsahují stejná data,
- žádné dva sloupce (atributy) nemají stejný název,
- hodnoty ve sloupcích by měly být atomické, dále nerozdělitelné.

# Identifikace záznamů – klíče

- Záznamy (řádky tabulek) potřebujeme nějak identifikovat, abychom je mohli spojovat, vybírat je z databáze apd.
- **Klíč** – množina atributů jednoznačně identifikující řádek tabulky. Těchto množin může být mnoho.
- **Kandidátní klíč** – nejmenší možná množina atributů, která jednoznačně identifikuje řádek tabulky. Může existovat více takových množin.
- **Primární klíč** – jeden vybraný kandidátní klíč, skutečně použitý pro identifikaci řádků tabulky (např. číslo na psí známce). Každá tabulka má jediný primární klíč, kterým jsou jednoznačně identifikovány řádky tabulky.
- **Cizí klíč** – množina atributů v tabulce, která je v jiné tabulce primárním klíčem (pomocí cizího klíče se odkazujeme z dané tabulky na záznam v jiné tabulce).



# Propojení tabulek pomocí klíčů



# Operace relační algebry

- DBMS vykonává na relacích následující operace, aby dovedl vrátit požadovaná data z databáze:
  - Množinové operace s relacemi: Sjednocení, průnik, rozdíl, kartézský součin (složení),
  - projekce – výběr pouze určitých atributů,
  - restrikce – omezení řádků relace pomocí operátorů  $<$ ,  $<=$ ,  $>$ ,  $>=$ ,  $=$ ,  $<>$ ,
  - spojení – spojení tabulek přes vybrané atributy (cizí klíče),
  - dělení.